INTRINSICALLY DISORDERED PROTEINS MAY SELECT PARTNERS BY FOLD

A Senior Scholars Thesis

by

KIM LANI GONZALEZ

Submitted to the Office of Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for the designation as

UNDERGRADUATE RESEARCH SCHOLAR

December 2010

Major: Genetics

INTRINSICALLY DISORDERED PROTEINS MAY SELECT PARTNERS BY FOLD

A Senior Scholars Thesis

by

KIM LANI GONZALEZ

Submitted to the Office of Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for the designation as

UNDERGRADUATE RESEARCH SCHOLAR

Approved by:

Research Advisor:	Sarah Bondos
Director for Honors and Undergraduate Research:	Sumana Datta

December 2010

Major: Genetics

ABSTRACT

Intrinsically Disordered Proteins May Select Partners by Fold. (December 2010)

Kim Lani Gonzalez
Department of Biochemistry and Biophysics
Texas A&M University

Research Advisor: Dr. Sarah Bondos
Department of Molecular and Cellular Medicine

Intrinsically disordered proteins lack a rigid structure due to their simple amino acid sequence. Because of their multiple roles, disordered proteins often account for a majority of proteins known to be associated with various diseases. In particular, these unstructured proteins tend to have pivotal roles in protein-protein interactions that occur in cell signaling. For structured proteins, surface topology complementarities play a crucial role in the selection of protein interactions and stabilization of the interaction interface. In contrast, the mechanism through which intrinsically disordered proteins select binding partners still remains unresolved. To gain more knowledge of binding interactions of disordered proteins, we examined the binding tendencies of Ultrabithorax, Ubx, a Hox protein consisting of both structured and intrinsically disordered regions. Of the 33 known proteins bound by Ubx, 15 have only 3 of a possible 1200 folds, suggesting Ubx selects protein partners by topology. To determine whether the structured or disordered region mediate binding, and thus topological selection, we used a yeast two-hybrid system to evaluate the strength of interactions

between Ubx and its binding partner. Results indicate that the intrinsically disordered regions of this protein are the reason for binding its partner. This is the first demonstration that surface topology is an important criterion for protein interactions involving intrinsically disordered regions.

DEDICATION

This thesis is in dedication to my mother, Blanca Gonzalez, who told me to always strive to do my best. On my sixteenth birthday, she told me in regards to my future, “I hope you dance.” This dance is for you.

ACKNOWLEDGMENTS

First of all, I would like to thank Texas A&M University for giving me an incomparable undergraduate education and a multitude of opportunities. I would also like to thank Dr. Sarah Bondos for allowing me to work in her lab and advising me on research and graduate schools. Thanks also go out to Hao-Ching Hsiao for aiding me in my lab work.

NOMENCLATURE

SCOP	Structural Classification of Proteins
SD	Standard Dropout
ONPG	o-nitrophenyl β-D-galactopyranoside
Ubx	Ultrabithorax

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

The amino acid sequence is the primary structure of all proteins. Interactions between these amino acids, like hydrogen or disulfide bonds, give rise to a unique protein fold. When binding a substrate or partner protein, structured proteins are typically expected to abide by induced fit or the lock and key mechanism. In both of these mechanisms, the structure of the interacting proteins is a critical determinant of the ability to bind. The need for proteins to have a certain shape in order to bind a partner or substrate is not the only requirement; the interacting proteins must have chemical compatibility. For instance, if two proteins are negatively charged throughout, they will not create an intimate contact regardless of their structure.

A complication with this rule arises when the protein lacks structure. Intrinsically disordered proteins lack a rigid structure due to a simple amino acid sequence, generally enriched in charged amino acids and depleted in hydrophobic residues (1). Intrinsically disordered regions often engage in protein-protein interactions, allowing a protein to bind to its partners with high specificity, yet can be easily reversed without many structural requirements (1). In graph theory parlance, “hubs” or proteins that engage in multiple interactions are more likely to be disordered. Little is known about the rules

---

This thesis follows the style of *Journal of Biological Chemistry*.

governing these unstructured proteins.

Disordered proteins in nature

Intrinsically disordered proteins occur in all three kingdoms, but are more prevalent in eukaryotes (2). Eukaryotes had 35-51% disordered proteins in its proteome in a study spanning 30 organisms, in comparison to bacteria having 6-33% and archaea having 9-37% disordered proteins (3). In 2006, Haynes *et al* compared the amounts of intrinsically disordered proteins in humans, *C. elegans*, *D. melanogaster*, and *S. cerevisiea* and found that humans had the most intrinsic disorder compared to the other three eukaryotes. The high demand for cell signaling in these organisms may account for the increase in intrinsically disordered proteins (3). Indeed, disordered proteins are known to play major roles in regulation, transcription, and translation (4). These unstructured proteins also have key functions in protein-protein interactions that occur in cell signaling (5). Problems with the protein-protein interactions could ultimately trigger disease. Disordered proteins account for 79% of proteins and 66% of signaling proteins known to be associated with cancer (2). In addition to having association with cancer, about 57% of cardiovascular proteins are also intrinsically disordered (6).

Because intrinsically disordered proteins participate in various vital functions in many organisms, it is imperative to understand their protein interaction regulations. Because these unstructured proteins appear in multiple diseases, rather than a single disease, this research is applicable to more fields of study. If we can better understand these

interactions, we would not only gain insight into the normal function of an organism, but we could also pave the way for the development of therapies or medicines for diseases caused by protein-protein interactions.

Single protein, multiple interactions

Little is known about the mechanism through which intrinsically disordered proteins select binding partners. Disordered proteins tend to be a part of biological networks, acting as "hubs" for protein-protein interactions (1). These signaling hubs can have interactions ranging from a few hundred or a few millions, and removal of hub proteins is frequently fatal (1). A common opinion is that the intrinsic disorder could be an advantage in regards to binding; the flexible unstructured region of a protein could allow binding to multiple partner conformations (1). Intrinsic disorder permits these hub proteins to partake in a plethora of interactions; they do not need to follow the induced fit or lock and key mechanisms (1). But what is allowing an intrinsically disordered protein to recognize its partner, and how can it bind *specifically* to many different partners?

Tompa and Fuxreiter proposed that the proteins might be disordered when unbound, but may gain some structure upon binding its partners. They proposed the flexibility could be due to three different models: the clamp model, the flanking model, or the random model. The clamp model looks at proteins that form structured ends separated by a loop of disorder when binding to another protein (7). The flanking model suggests the

opposite; the structure arises in the middle of the protein when bound, but has tails of intrinsic disorder on either end (7). The random model, the most extreme of the three, insinuates that the regions of disorder stay unstructured when binding (7). These models focus on the importance of structure, a major aspect for protein function, with regards to binding. We are also interested in how this disorder contributes to binding and chose to examine protein-protein interactions of an intrinsically disordered protein.

Ultrabithorax is a disordered protein

Ultrabithorax, Ubx, is an intrinsically disordered Hox protein consisting of both structured and intrinsically disordered regions (8). There are four regions of disorder in Ubx, but the first two being less than 20 amino acids in length, would not be a significant contributor to binding interactions (8). For our purposes, the term "disordered region" will refer to a segment having more than twenty amino acids.

Previous research showed that Ubx interacts with 29 partner proteins. Within this group the partners shared only five folds. Given there are 1182 known protein folds found in the Structural Classification of Proteins, SCOP, Ubx appears to be selecting partners by topology. The structure of the Ubx sequence is mainly intrinsically disordered in amino acids 1-216 and in the microexon region N-terminal to the homeodomain (8). Ubx also has structured regions, mainly the homeodomain, that may contribute to partner selection (8). If the disordered regions determined partner selection, we could use Ubx to ascertain the rules for the interactions by unstructured proteins. Preliminary results suggest that

the binding partner is selected by the disorder. Since little is known about unstructured protein binding properties, we will also determine whether intrinsically disordered proteins generally select binding partners by fold.

We have several reasons to investigate this subject. First, we would like to test the role of the intrinsic disorder portion of Ubx in protein-protein interactions. If the intrinsically disordered regions are required for binding, we could confirm that Ubx selects its binding partner by shape. Second, if this can be proven, the next test would be to determine if other intrinsically disordered proteins select binding partners by fold as well. Third, by inspecting known interactions, we must determine if the selection is seeking grooves or a canyon-like shape, which increases contact or interactions with the binding partners.

In this work we use yeast two hybrid to investigate our hypothesis. LexA is a DNA binding domain that we fuse to Ubx, whereas the binding partner has a B42 activating domain attached. If the binding partner and Ubx interact, then the B42 activating domain and LexA DNA binding domain are in a single protein complex that can activate the transcription of genes downstream of LexA DNA binding sites. However, since the Ubx protein already contains an activation domain, we have to alter one of two things to prevent self-activation: sever the N-term to the 216th amino acid or utilize a proline mutant. The central question is whether or not the binding partners need the intrinsically disordered regions to bind. Therefore, we also examined binding of Ubx with the

intrinsic disorder regions- the N-term to the 216th amino acid and the microexon region removed. If the partner does not bind without both regions of disorder, we will test if either the N-term to the 216th amino acid region or the microexons can restore binding. β-galactosidase liquid assays will be employed in testing the protein-protein interactions to quantitatively compare binding to Ubx and its variants.

CHAPTER II

EXPERIMENTAL PROCEDURES

Yeast two-hybrid screen

The reporter host strain is *Saccharomyces cerevisiae* EGY48 [MATR ura3 his3 trp1 LexAop(x6) LEU2]; this strain also carries a wild-type LEU2 gene directed by a series of LexA operators. A reporter plasmid, p8op-lacZ, transformed EGY48 and carries the lacZ reporter gene. This gene is also controlled by LexA operators in order to ensure stable transformation before testing activation by introduction by plasmids.

This yeast was grown and then inoculated in Glu/-Trp/-Ura liquid media overnight in a 30°C at 225 rpm. After removing from the incubator, 500 μl of the yeast was pipetted into 1.5 ml Eppendorf tubes, which are labeled for the desired DNA to be transformed. These tubes were centrifuged for 30 seconds at 10,000 rcf. The supernatant was then removed and the cell pellet was washed sequentially with two different buffers. Then 500 ng of DNA to be transformed was added to the washed cells. 10 μl of the salmon sperm was added next. The 10 mg/ml salmon sperm DNA was denatured in an Eppendorf thermomixer at 96°C and 300 rpm for six minutes. The 1.5 ml Eppendorf tube was then vortexed at a medium speed for 10 seconds. These tubes were then placed into a hot water bath for one hour at 45°C. After incubation the tubes were centrifuged for 20 seconds on the mini-centrifuge, and the supernatant was discarded. Then about 120 μl of nuclease free water was added to the tube to resuspend the yeast cells using a

pipette. Aliquots of 40 μl of this suspension were dispensed onto three separate Gal/Raf/-His/-Ura/-Trp plates. These plates were then incubated for 4 to 5 days at 30°C.

β-galactosidase assays

The procedures for the β-galactosidase liquid assays were followed as outlined in the Clontech Yeast Protocols Handbook for liquid cultures using ONPG as a substrate. Overnight cultures in were prepared in liquid SD selection in test tubes. The fresh culture was incubated at 30°C while shaking at 230 rpm until the OD600 reading of 1ml of culture was between 0.5-0.8. This OD600 was recorded when the cells were ready to be harvested. Three 1.5 ml microcentrifuge tubes each received 1.5 ml of the harvested culture, and were centrifuged at 14,000 rpm for 30 seconds. The supernatant was then removed and the cells were resuspended in 1.5 ml of Z buffer, and centrifuged again at 14,000 rpm for 30 seconds. Afterwards, only 300 μl of Z buffer was used to resuspend the cells. Three new 1.5 ml microcentrifuge tubes each received 100 μl of the cell suspension, and then were stored at -80°C until needed.

The cells were removed from the freezer and heat shocked by submerging the tubes in liquid nitrogen for 45 seconds and then placing in 37°C water bath for 1 minute. This was repeated twice more. A blank, containing 100 μl of Z buffer, and the sample tubes each received 700 μl of Z buffer + β-mercaptoethanol and then 160 μl of fresh ONPG. The tubes were then placed in the 30°C incubator, and at this point the timer was set. When a yellow color developed 400 μl of stop buffer (1M Na_2CO_3) was added to the

tube, and the timer was stopped. These tubes were centrifuged for 10 minutes at 14,000rpm and about 1 ml of supernatant was transferred to clean, disposable cuvettes. The spectrophotometer was calibrated at λ_{abs}=420 nm using the prepared blank, and then the optical density measurements for the samples at 420 nm were recorded. Afterwards, the β-galactosidase units were calculated using the following equation:

$$\beta\text{-galactosidase units} = (1000 \times OD_{420})/(t \times V \times OD_{600})$$

Once all the β-galactosidase units were configured for every Ubx isoform, the data was then used to create the graphs seen in the results section.

During the process of growing and collecting samples, there was quite a labeling misfortune. These cultures had to be discarded, resulting in a few different Ubx isoforms not having as many various partners to be tested. During the research we attempted more samples than reported, but some samples did not provide colonies for the assay. In order to report data with validity, these samples were not reported in this paper. Of course these tests would provide more valid data, and are in the process of being completed.

CHAPTER III

RESULTS

Various isoforms of Ubx were tested to determine whether intrinsically disordered regions mediate binding to protein partners (Fig. 1). The regions of disorder lie between amino acids 103-216 and in the microexons (8). The role of the first disorder region, between amino acids 103-216, will be probed by examining a truncation and internal deletion. In contrast, the role of the microexons will be tested with naturally occurring mRNA splicing isoforms. The full-length isoform, pLexA-UbxIbPro4, contains all three of the unstructured microexons: b, I, and II (9). In comparison, the isoform pLexA-UbxIVaPro4 contains all regions except for the three microexons (9). We employed the pLexA-UbxIVaPro4 isoform to test the binding influences of the three microexons, while still maintaining the large segment of disorder at amino acids 103-216. This large segment of disorder at amino acids 103-216 is removed in the isoform pLexA-UbxIb103-216.

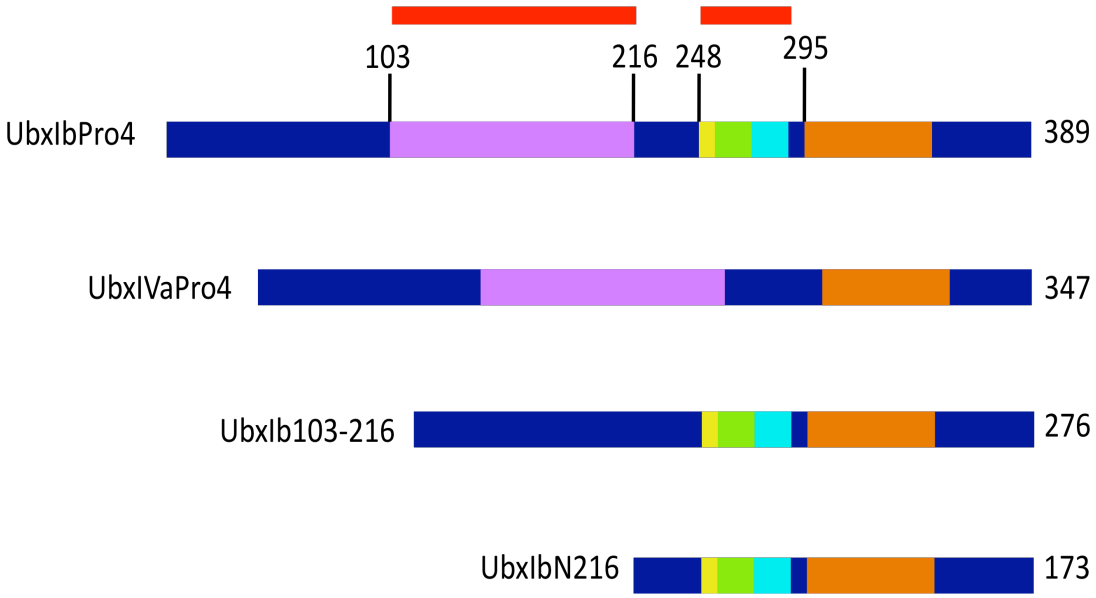

FIGURE 1. **Various Ubx isoforms used in yeast two-hybrid analysis.** This figure shows the four different isoforms of Ubx. The top figure, labeled UbxIb, is the full-length protein consisting of 389 amino acids. The red bars across the top denote intrinsically disordered regions over 20 amino acids in UbxIb. The purple section represents the disordered region between amino acids 103-216. The yellow segment represents the "b element" microexon, and is found at amino acid 248 in the full length Ubx isoform. The green segment represents microexon I. The light blue represents microexon II. The orange segment of the protein represents the homeodomain. The dark blue represents all other amino acids within the protein.

The partner used for these tests performed is CBP80, one of five Ubx partners with an alpha-alpha superhelix fold. All of the Ubx isoforms were fused to the pLexA binding domain and CBP80 has a B42 activating domain attached. If CBP80 and Ubx interact, then the B42 activating domain and LexA DNA binding domain form a single protein complex that can activate the transcription of genes downstream of LexA DNA binding sites. Nonetheless, the Ubx protein already contains an activation domain, therefore we have to remove the N-term to the 216th amino acid or utilize a proline mutant to prevent self-activation. Both isoforms containing "Pro4" in its name are proline mutants, and the other two isoforms remove the activation domain, and therefore do not require the proline mutation. Ubx and binding partner interactions activate the transcription of the

Lac-Z gene downstream of the pLexA DNA binding sites, resulting in β-galactosidase. The β-galactosidase liquid assays were employed to quantitatively assess the interactions between Ubx and its binding partner, measured in β-galactosidase units. The β-galactosidase units are indicative of binding interactions, the larger the β-galactosidase units the greater the binding interactions. The relative interactions between the Ubx isoforms and binding partner can be seen below (Table 1 and Fig. 2).

TABLE 1
Relative interactions between Ubx isoforms and binding partner CBP80
This table summarizes the average of the β-galactosidase liquid assays along with standard deviation for various binding partners among different Ubx isoforms. The highlighted gray area represents the controls for the media, where pB42AD-T+pLexA-BD-p53 is the positive control and pB42AD-T+pLexA-BD-Lam is the negative control. Each interaction was measured five times from five different transformations.

Ubx Isoform	β-galactosidase units
pLexA-UbxIbPro4	66.29 ± 26.69
pLexA-UbxIbN216	1.89 ± 0.04
pLexA-UbxIb103-216	39.49 ± 6.77
pLexA-UbxIVaPro4	24.89 ± 7.68
pB42AD-T+pLexA-BD-p53	171.29 ± 45.52
pB42AD-T+pLexA-BD-Lam	0.67 ± 0.30

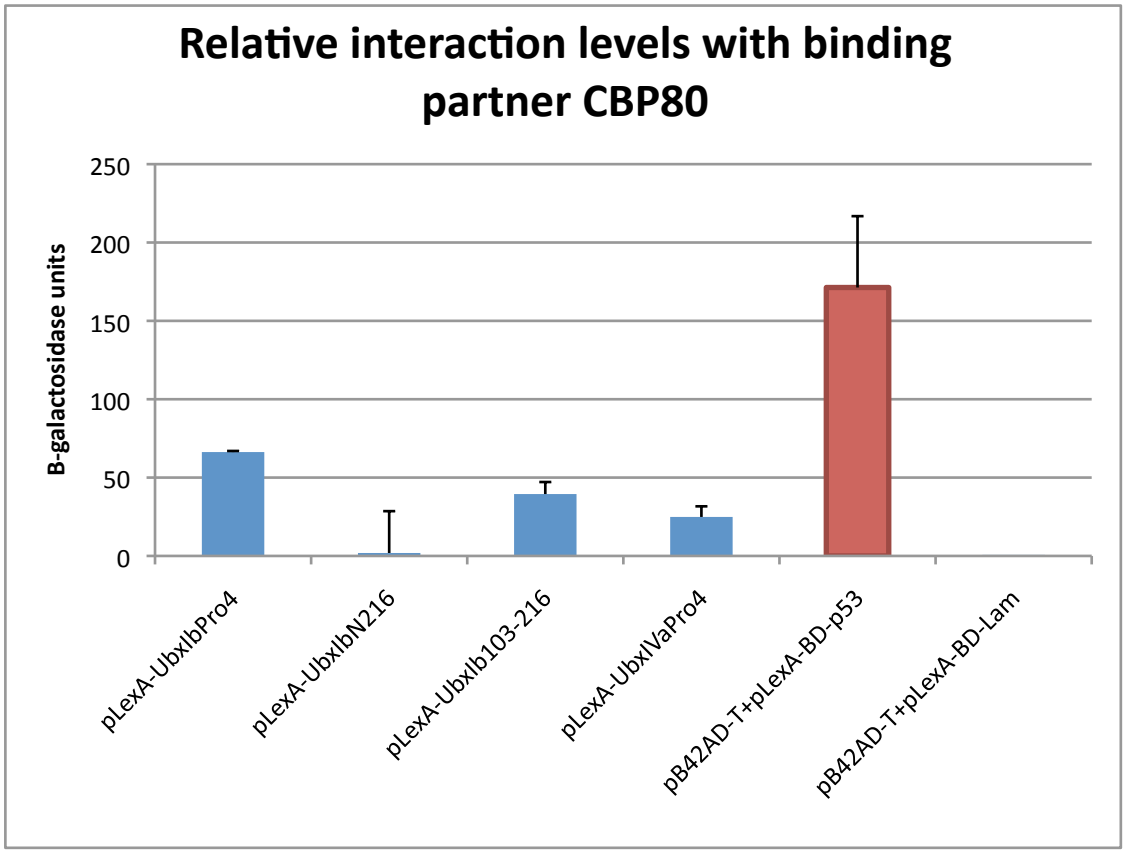

FIGURE 2. **Overall results from the β-galactosidase liquid assays.** This chart relays the average of the β-galactosidase liquid assays for various binding partners, as seen previously in Table 1. The first four bars to the left of the graph, in blue, denote the various Ubx isoforms that were tested using CBP80. The two red bars on the right of the graph represent controls for the media provided when the media was purchased.

The Ubx isoform pLexA-UbxIbPro4 had the strongest interaction with 66.29 ± 26.69 β-galactosidase units. With 1.89 ± 0.04 β-galactosidase units the Ubx isoform pLexA-UbxIbN216 had the weakest interaction with CBP80, a detectable level of interactions barely above background levels. The β-galactosidase units, and therefore binding interactions, markedly diminished when the region N-term to the 216th amino acid was removed. These results followed our expectations, and reinforced that the lack of intrinsic disorder does affect Ubx binding CPB80. The pLexA-UbxIbN216 isoform is

utilized to test the role of the intrinsic disorder within the region N-term to the 216th amino acid, and found this region contributes to binding. Since pLexA-UbxIbN216 does remove a large portion of the disorder within Ubx, it also removes a segment of structure at the N-terminus. This would consequently skew our findings; therefore we used pLexA-UbxIb103-216 to only remove the large portion of the disorder, but maintain the upstream structured region.

The other two isoforms revealed that the second unstructured region also play a role in protein-protein interactions. The isoform pLexA-UbxIb103-216 had 39.49 ± 6.77 β-galactosidase units, an increase in β-galactosidase units compared to the interactions made by pLexA-UbxIbN216. The isoform pLexA-UbxIVaPro4 had 24.89 ± 7.68 β-galactosidase units, also indicating an increase in binding interactions compared to pLexA-UbxIbN216. When comparing the two interaction levels of pLexA-UbxIb103-216 and pLexA-UbxIVaPro4, the results do not concur with our hypothesis. Because pLexA-UbxIb103-216 has a larger portion of intrinsic disorder removed, one would think that pLexA-UbxIVaPro4 should have slightly more β-galactosidase units than pLexA-UbxIb103-216. There are three possible explanations for this result. First, the structured regions N-term to the 103rd amino acid could have an influence on binding, thus the minimal decrease in β-galactosidase units when the amino acids 103-216 were removed. Second, there are small regions of disorder of less than 20 amino acids N-term to the 103rd amino acid, which could also affect binding, and lessen the severity of

loosing the major region of disorder. Third, these results can be due to residual transcription activation by Ubx, since the proline mutation is absent.

CHAPTER IV

SUMMARY AND DISCUSSION

Summary

When examining the effect microexons have on binding, by comparing pLexA-UbxIbPro4 and pLexA-UbxIVaPro4, it is apparent that the microexons do contribute to CBP80 binding. The Ubx N216 truncation shows almost no interaction between Ubx and CBP80, deeming this region a major determinant of CBP80 binding. Looking at the two interaction levels of pLexA-UbxIb103-216 and pLexA-UbxIVaPro4, it seems as though the microexons may have a larger effect on binding interaction than the large portion of disorder from amino acids 103-216. Comparing pLexA-UbxIbN216 and pLexA-UbxIb103-216 indicates that either the mostly structured region N-term to the 103rd amino acid also impacts binding or pLexA-UbxIb103-216 has higher baseline levels of auto-activation.

Discussion

Further examinations of the microexons are to be conducted to evaluate the effect of each region on Ubx binding its partner protein. The varying levels of protein-protein interactions as the result of splicing could be an advantage in vivo. The use of microexons for protein interactions could allow a developing animal to regulate protein interactions through stage and tissue specific alternative splicing (10).

The graduate student working on the same project, Hao-Ching Hsiao, has recently carried out experiments using Ubx isoforms and multiple binding partners that were not discussed in this paper. The patterns observed for CBP80 interactions are also observed for the other partners as well, with the only exception being that the isoform pLexA-UbxIb103-216 has more β-galactosidase units than pLexA-UbxIVaPro4. The trends are similar for other partners with an alpha-alpha super helix fold and partners with different folds. This reinforces the thought that the intrinsic disorder selects partners by topology.

The hypothesis that the intrinsic disorder would prefer a protein partner with grooves or a canyon-like shape is still being tested. The canyon-like shape would surround the intrinsically disordered region, increasing number of molecular interactions with the binding partners. This increase in interactions would likely lead to a higher expression of β-galactosidase units when comparing to a partner that has one tangent interaction between the disordered region of Ubx and its partner. This presumption concurs with the initial results of testing done in Indiana by A. Keith Dunker that suggest that the partners with canyon-like folds bind more intrinsically disordered proteins than partners without this canyon like fold.

REFERENCES

1. Dunker, A. K., Cortese M. S., Romero, P., Iakoucheva, L. M., and Uversky, V. N. (2005). *FEBS J*. **272**, 5129–5148

2. Oldfield, C. J., Cheng, Y., Cortese M. S., Romero, P., Uversky, V. N., and Dunker, A. K. (2005). *Biochemistry* **44**, 12454-12470

3. Dunker, A. K., Brown, C. J., Lawson, J. D, Iakoucheva, L. M., and Obradovic, Z. (2002). *Biochemistry* **41**, 6573-6582

4. Shimizu, K., and Toh, H. (2009). *J. Mol. Biol*. **392**, 1253–1265

5. Haynes, C., Oldfield, C.J., Ji, F., Kiltgord, N., Cusick, M.E., Uversky, V.N., Vidal, M., and Iakoucheva, L.M. (2006). *PLoS Comput. Biol*. **2**, 0890-0900

6. Cheng, Y., LeGall, T., Oldfield, C. J., Dunker, A. K., and Uversky, V. N. (2006) *Biochemistry* **45**, 10448-10460

7. Tompa, P., and Fuxreiter, M. (2007). *Trends Biochem. Sci.* **33**, 2-8

8. Liu, Y., Matthews, K. S., and Bondos, S. E. (2008). *J. Biol. Chem*. **283**, 20874-20887

9. Liu, Y., Matthews, K. S., and Bondos, S. E. (2009). *J. Mol. Biol.* **390**, 760-774

10. Tan, X. X., Bondos, S. E., and Matthews, K. S. (2006). *Mol. Cell. Proteomics* **5**, 824-834

CONTACT INFORMATION

Name:	Kim Lani Gonzalez
Professional Address:	c/o Dr. Sarah Bondos Department of Molecular and Cellular Medicine MS 1114 Texas A&M University College Station, TX 77843
Email Address:	kimgonzalez11@gmail.com
Education:	B.S., Genetics, Texas A&M University, December 2010 Cum Laude Undergraduate Research Scholar